Tax Evasion as a Global Game (TEGG) in the laboratory*

Miguel Sanchez Villalba
University of Alicante

April 2007

Abstract

Sanchez Villalba, 2007) claims tax evasion can be modelled as a global game when income shocks are common and prescribes that the tax agency should audit each individual taxpayer with a probability that is a non-decreasing function of every other taxpayer's declarations ("contingent policy rule").

This paper uses experimental data to test the predictions of the model and finds supporting evidence for the hypothesis that the contingent policy rule is superior to the alternative "cut-off" one.

It also finds that data fits the qualitative predictions of the global game model, regarding both participants' decisions and the experiment's comparative statics.

JEL Classification:

Keywords: Tax Evasion, Global Games, Experimental Economics, Rationality, Information, Beliefs

*I thank Frank Cowell, Georg Weizsacker and seminar participants at the London School of Economics for helpful comments and discussions. I gratefully acknowledge financial support from proyecto SEJ 2007-62656.

1 Introduction

Common income shocks that affect rather homogeneous agents in similar ways are well documented: the fact that airlines' sales plummeted after 9/11, chicken breeders faced low demand after the avian flu outbreak, and emergent markets have difficulties attracting investors every time the U. S. Federal Reserve increases interest rates are just a few examples one can bring forward. Furthermore, they show that often these shocks are the main source of income variability, with a *common/idiosyncratic* ratio well above 1.

It is therefore not surprising that a tax agency that ignores them when deciding its auditing policy will choose a clearly suboptimal one. But this is exactly what happens if they follow the most popular policy prescribed by the literature: the "cut-off rule" (Reinganum and Wilde, 1988)). It states that the agency should not audit any firm that declares about or above a certain fixed cut-off income level, while auditing those who declare below it with a sufficiently high probability. Combined with common income shocks, this policy leads to systematic mistargeting: the agency audits "too much" in bad years and "too little" in good ones.

In this scenario, Sanchez Villalba, 2007) finds that the optimal policy (named "contingent rule" by the author) consists of the agency auditing every firm with a probability that is a non-decreasing function of every other taxpayer's declarations. This is because other firms' declarations give the agency information about the realisation of the shock and so the probability of a given taxpayer being an evader is (weakly) higher the higher are her fellow taxpayers' declarations.

The purpose of this paper is therefore to test Sanchez Villalba's model (henceforth, TEGG model, or "Tax Evasion as a Global Game" model). This is a relevant task because it will help determining which of the alternative rules (contingent or cut-off) is superior to the other and, indirectly, whether the data is consistent with the modelling of tax evasion as a global game and its associated predictions.

However, real-world data on tax evasion is not readily available. Those who engage in tax evasion are not willing to indicate it for obvious reasons, but even tax agencies are reluctant to provide the information because of the confidentiality of tax returns: even if the datapoints are not labelled, in many cases it is quite easy to identify which individual firm they belong to, thus revealing sensitive information that could affect the company negatively.

For this reason, the current paper will use the second-best available dataset, namely, the one collected in a computerised experiment in which participants interacted with each other in situations that resembled the scenario described by the TEGG model. This methodology has the obvious disadvantage of making difficult the extrapolation of results from the sample to the population, but allows the experimenter a greater control over the variables under study and is, as mentioned before, the only available one anyway.

The econometric analysis finds that the agency is better off when using the contingent rule than when using the cut-off one, and so that the key prediction of the TEGG model is strongly supported. It also supports the hypothesis that people make decisions (qualitatively) consistent with higher-order beliefs (which play an important role in ensuring the uniqueness of the global game equilibrium) and that the comparative statics follow the ones predicted by the global game method.

To the best of my knowledge, nobody tested empirically (using either real-world or experimental data) the predictions of a TEGG-like model, but plenty of laboratory experiments were framed as/based on tax compliance problems. The closest reference is Alm and McKee, 2004), which analyses tax evasion as a coordination game. In contrast, the present analysis considers it as a global game, which requires not only the strategic uncertainty generated by the coordination game but also the "fundamental uncertainty" created by the incompleteness of information regarding the payoff functions. Tests of the global game technique seem to support it in terms of predictive power (Cabrales *et al.*, 2002)) and/or comparative statics (Heinemann *et al.*, 2004b)), but are less supportive of the participants' use of "higher-order beliefs" when making decisions. The latter result is also the conclusion of other studies, like Stahl and Wilson, 1994) and Bosch-Domenech *et al.*, 2002).

2 Tax Evasion as a Global Game

The global game methodology (Carlsson and van Damme, 1993), Morris and Shin, 2002)) is a mechanism that, thanks to the existence of some uncertainty about the payoff functions of the players, selects one of the multiple equilibria of a coordination game.

Sanchez Villalba, 2007) claims that, in the presence of common income shocks, tax evasion can be modelled as a global game because the agency's optimal policy generates a coordination game and taxpayers' imperfect information about the agency's "type" creates the uncertainty about payoffs.

Drawing on the fact that most tax agencies worldwide partition the population of taxpayers into categories where members share some non-manipulable characteristics, he analyses the agency's problem within each one of them. The high degree of homogeneity within a category implies that the idiosyncratic shocks will be small compared to the common ones, and so, for all practical purposes, one can assume that every member has the same income y: it is high ($y = 1$) in "good" years (which occur with probability γ) and low ($y = 0$) in "bad" ones (with probability $1 - \gamma$).

The timing of the game is such that in the first stage all actors learn their private information, then taxpayers submit their declarations, and finally the agency (after observing all declarations) undertakes audits (if any). An agency's private information is its "type", parameterised by $\lambda \in [0, 1]$ and interpreted as the loss associated with letting an evader get away with her evasion. In turn, the private information of a taxpayer i consists of her income $y_i \in \{0, 1\}$ and her signal $s_i := \lambda + \varepsilon_i$, where ε_i is a white noise error term. This signal embodies all the information about the type of agency available to the taxpayer (news, previous experience, conversations with colleagues/friends, etc.). All actors (taxpayers and agency) know every parameter of the game and their own private information. They also know the probability distributions of other actors' private information, but not their realisations.[1]

Every taxpayer has to decide how much income to declare, d, in order to maximise her expected utility. The optimal declaration strategy follows the standard literature except for the fact that, since the exact probability of detection a is unknown to the taxpayer, her declaration will be a (weakly) increasing function of her **belief** about a.

The agency chooses a in order to minimise the losses associated with making targeting errors. These can take two forms: *zeal* errors (Z) occur when resources are wasted on auditing compliant taxpayers; *negligence* errors (N) take place when evaders are not caught and the corresponding fines are not collected.[2] The agency minimises a "loss function" that aggregates errors into one metric and can be written as $L = \lambda N + (1 - \lambda) Z$, where λ is the above mentioned (privately known) agency's type.

Sanchez Villalba's new insight is that the agency's optimal auditing policy regarding taxpayer i, a_i, is (weakly) increasing in the agency's type, λ, and the declarations of every other taxpayer in the category, d_j, $j \neq i$. The last result is especially important because it generates a *negative externality* between taxpayers: the higher the declaration of a taxpayer j, the higher the probability that another taxpayer i ($i \neq j$) is audited and the lower the latter's expected utility. Together with the optimal declaration strategy, this creates the *strategic complementarities* between taxpayers' declarations that constitute the defining feature of a *coordination game*. Specifically, the higher the declaration of taxpayer j, the higher the incentives of taxpayer i to comply as well.

The associated problems of multiplicity of equilibria are, however, side-stepped because of the taxpayers' uncertainty about the realised agency's type, λ, and the heterogeneous beliefs about a they derive from their disparate private signals, $E(a|s)$. This "fundamental uncertainty", plus the "strategic uncertainty" generated by the coordination game, create the conditions for modelling tax evasion as a global game. This leads, through a process

[1] Except in the case of income y, of course, because it is assumed that everyone in a category has exactly the same level of income. Adding some income heterogeneity avoids this "perfect observability" issue but does not provide any new insight or affect the predictions, so for simplicity this avenue is not pursued.

[2] Formally, a zeal error ($Z = 1$) occurs when $a(1 - (1 - d) y) = 1$; a negligence one ($N = 1$) when $(1 - a)(1 - d) y = 1$. Implicit in the latter formula is the assumption that it is always profitable for the agency to audit a known evader (i.e., in such cases the fine is greater than the cost of the audit). The alternative possibility implies the uninteresting solution where nobody is audited, even known evaders.

akin to the "iterated deletion of strictly dominated strategies" (IDSDS) method, to a unique equilibrium: in each iteration, signals provide information about what other taxpayers will *not* do, and in the end it ensures that only one strategy survives, namely, one where taxpayers with low signals (and hence low beliefs about being discovered) evade, while those with high signals comply. Furthermore, equilibria with full, partial and zero evasion can arise, depending on the value of the parameters.

The key prediction of the TEGG model is that an agency that implements the "contingent" policy will do (weakly) better than if it implemented the standard "cut-off" one, *ceteris paribus*. Testing this hypothesis is the main purpose of the present study, though the experimental dataset is rich enough as to allow for the investigation of others that will also be analysed, like the use of higher-order beliefs or the comparative statics generated by changes in the parameters of the problem.

3 Experiment design

The experiment took place on the 28 of November 2006, at the ELSE computer laboratory (UCL, London). The pool of participants was recruited by ELSE from their database of about 1,000 people (most of them UCL students). 200 of them were chosen randomly and invited to take part and the first 100 who accepted the offer were allocated to sessions according to their time preferences.[3] No person was allowed to participate in more than one session.

The day of the experiment 76 people took part in four treatments (labelled GC, GE, LC and LE for reasons to be explained later in this section), each involving a 60-to-90-minute long session. Participants were lined up outside the lab according to their arrival time. At the designated time they entered and freely chose where to sit. They were not allowed to communicate for the entirety of the session and could not see other people's screens.

Each session consisted of 6 sections, namely, instructions, short quiz, trial rounds, experimental rounds, questionnaire and payment. The instructions were read aloud by the instructor and, in order to ensure their correct understanding, the participants were asked to complete a "short quiz" (shown in appendix A; correct answers and the rationale for them were provided by the instructor after a few minutes). For the same reasons, participants then played two "trial" (practice) rounds whose outcomes did not affect their earnings. After each of these first three stages the instructor answered subjects' questions in private. The experimental rounds (20 per session) were then played, and after that, subjects completed

[3] Five "reserve" people were invited to each session and 7 of them had to be turned down because the target number (20 per session) was reached or because the treatment required an even number of participants (treatments GC and GE). Each one of them was paid the £5 show-up fee before being dismissed.

a questionnaire with information regarding personal data and the decision-making process they followed. Finally, each participant was paid an amount of money consisting of a fixed show-up fee (£5) and a variable component equal to the earnings accumulated over the 20 experimental rounds.[4] Table 1 shows the exchange rate used to translate experimental currency into money, as well as other payment-related summary statistics.[5]

Treatment	Participants	£ per 1000 points	Min/Avg/Max Payment		
GC	18	0.50	10.80	11.52	11.80
GE	18	0.90	7.40	9.30	9.80
LC	20	0.50	11.60	11.65	11.80
LE	20	0.90	9.80	11.20	11.60
All	76		7.40	10.95	11.80

Note: £ per 1000 points is the exchange rate at which 1000 "experimental points" where transformed into pounds.

Table 1: Treatments. Participants and Money.

Each experimental round consisted of two stages: the "Choice" one, where participants had to make a decision that would affect their payoffs, and the "Feedback" one, where they got information about the round outcome.

		Column player	
		Y	Z
Row player	Y	$x(Y, Y, q)$	$x(Y, Z, q)$
	Z	$x(Z, Y, q)$	$x(Z, Z, q)$

Note: Only Row player's payoffs (x) are shown. Payoff's components are Row's action, Column's action and the realisation of the random variable q. Column's payoffs are symmetrical.

Table 2: Stage game.

In the "choice" stage a one-shot game was played where the subjects had to choose one of two possible actions (Y or Z) interpreted as *Evasion* and *Compliance*, respectively (the game's normal form for the 2-person case is shown in table 2). A participant i's payoff was determined by her own decision, $d_i \in \mathcal{D} := \{Y, Z\}$, the decisions of the other $n-1$ people in her category, $\mathbf{d}_{-i} := (d_1, ..., d_{i-1}, d_{i+1}, ..., d_n)$, $\mathbf{d}_{-i} \in \mathcal{D}^{n-1}$, and the realisation of a random variable, $q \in \mathcal{Q} := \{A, B, C\}$. Formally,

$$x_i := x(d_i, \mathbf{d}_{-i}, q) \tag{1}$$

Different choices have different effects on payoffs, and so, while the payoff from option Y

[4] In other experimental studies (Heinemann *et al.*, 2004a) among them) participants were paid according to the result of one randomly-chosen round. The rationale for this is that it avoids hedging, something that is not a problem here: the maximum payment a person can receive in any given round is £0.50 or £0.90 (depending on the treatment), with expected values in the £0.30-£0.35 range.

[5] In order to minimise delays and computational hassle, every person's payment was rounded *up* to the closest multiple of £0.20. Participants were not told about this arrangement until after they completed their questionnaires in order to avoid strategic play with respect to this peripheral matter.

is uncertain (reflecting the uncertainty about being audited), that of option Z is a known, fixed quantity: formally, for every $\mathbf{d}_{-i}, \mathbf{d}'_{-i} \in \mathcal{D}^{n-1}$, $q, q' \in \mathcal{Q}$,

$$x(Z) := x(Z, \mathbf{d}_{-i}, q) = x\left(Z, \mathbf{d}'_{-i}, q'\right) \tag{2}$$

The random variable q can take values A, B and C with probabilities $p(A) = 0.20$, $p(B) = 0.60$ and $p(C) = 0.20$, respectively. It represents the different possible "types" of agency (A : soft (on evasion), B : medium, C : tough) and corresponds to the "λ" mentioned in the previous section. It affects evasion payoffs (i.e., Y-payoffs) negatively: the tougher the agency, the more likely the evader will be audited and the lower her payoff. Formally, for every $\mathbf{d}_{-i} \in \mathcal{D}^{n-1}$,

$$x(Y, \mathbf{d}_{-i}, A) > x(Y, \mathbf{d}_{-i}, B) > x(Y, \mathbf{d}_{-i}, C) \tag{3}$$

At the time of making a decision participants do not know the value of q, but each one of them gets a private signal $s \in \mathcal{S} := \{a, b, c\}$ (called "hint" in the experiment) that is related to the realised value of q as shown in table 3 (and in the Instructions sample in appendix A). The instructions highlighted the fact that different people could get different hints but q was the same for everyone. No other probabilities were provided explicitly, though the instructions did supply the information required for their computation, namely, the prior probability distribution of q, $p(q)$, and the conditional one, $f(q|s)$.[6]

If hint =...	...then q =...	...with probability $f(q\|s)$ =...
a	A	1.000
	A	0.125
b	B	0.750
	C	0.125
c	C	1.000

Table 3: Hints and q.

The participant's submission of her decision (Y or Z) ended the "Choice" stage and gave way to the "Feedback" one, in which the person was informed about the realised value of q, the signal she received, her choice and her payoff for the round. At no stage was a subject given any information about the signals or choices of any other participant.[7]

By clicking on the "Continue" button, participants exited the "Feedback" stage and moved on to the next round (if any was left). Rounds were identical to each other in terms of their structure (Choice and Feedback stages) and rules (payoff computations, prior and conditional probability distribution of q), but may have differed in the *realised* values of the

[6] A "Choice stage" screenshot (labelled "Choice screen" in the experiment) can be seen in the instructions sample in appendix A. The programme used was z-Tree (Fischbacher, 2007)).

[7] A "Feedback stage" screenshot (labelled "Results screen" in the experiment) can be seen in the instructions sample in appendix A.

random variables (q and $\mathbf{s}$). Participants were told explicitly about this and informed that each round was independent from every other one.

3.1 Treatments

The experiment's treatments were defined according to the policy used (contingent v cut-off, or "global" (G) v "lottery" (L)) and the predicted optimal strategy of the participants (which for this experiment, as will be shown later, reduces to determining the optimal choice when hint b is received: to evade (E) or to comply (C)).[8] This way the experimental setup can be visualised as in table 4.

		Participant's strategy	
		Comply (C)	Evade (E)
Auditing rule	Contingent (G)	GC	GE
	Cut-off (L)	LC	LE

Note: Participant's strategy refers to the optimal strategy of a participant when receiving hint b.

Table 4: Treatments.

The difference between Global and Lottery treatments is related to the effect of other subjects' choices on the payoffs of individual participants. In the Lottery treatments the rule implemented by the agency is of the cut-off type, and so what other people do does not affect taxpayer i's payoff. Formally, for every $q \in \mathcal{Q}$,[9]

$$x\,(Y, Y, q) = x\,(Y, Z, q) \tag{4}$$

In Global treatments, on the other hand, the auditing policy followed is the contingent one, implying that other people's declarations have a negative impact on taxpayer i's payoff via the increased probability of detection. Formally, for every $q \in \mathcal{Q}$,

$$x\,(Y, Y, q) > x\,(Y, Z, q) \tag{5}$$

[8] Tax evasion has often been compared to a gamble in which the taxpayer "wins" (i.e., gets away with evasion) with probability w, and "loses" (i.e., is caught and has to pay a fine on top of the unpaid taxes) with probability $1 - w$.

The cut-off rule is equivalent to a standard lottery (and hence the name of the treatment) because it *fixes* the chances of winning (say $w = 1 - p$) and losing ($1 - w = p$). Evasion can therefore be seen as equivalent to buying $(1 - p)\,N$ out of a total pool of N raffle tickets, each one of them equally likely to be the winner.

In the Global treatments, on the other hand, those probabilities are *not* fixed, because they are affected by what other people do. In particular, since other people's compliance has a negative impact on my payoff, the fact that other people comply is equivalent to having the total number of tickets increased to, say, $N' > N$, so that my probability of winning w' (in spite of my holding the same number of tickets as before, $(1 - p)\,N$) is now comparatively lower: $w' = \frac{(1-p)N}{N'} < \frac{(1-p)N}{N} = w$.

[9] I restrict my attention to the 2-person case, which will be the relevant one throughout the paper. The extension to the n-person case is straightforward.

It is worth mentioning here that the Lottery treatment can be interpreted as a special (limit) case of the Global one in which the effect of other people's decisions on a certain participant's payoff is arbitrarily small. Consequently, and without loss of generality, henceforth the analysis will be restricted to the Global case, with the occasional reference to the Lottery one provided only when relevant.

For the experiment, participants in the Global treatments were divided in 9 groups of 2 people each, the matching protocol being random (equi-probable) within rounds and independent across them. The experimental setup reproduced the three typical scenarios described by the global game literature:

† The two **extreme** cases in which the "fundamentals" are "so bad"/"so good" that there is a strictly dominant strategy which is chosen by everyone. In the experiment the fundamental is q, the agency's "toughness", and so strict dominance requires that everyone should evade when the agency is very soft ($q = A$) and that everyone should comply when it is very tough ($q = C$). Formally, for every $d' \in \mathcal{D}$,

$$x(Y, d', A) > x(Z) \quad (6)$$

$$x(Y, d', C) < x(Z) \quad (7)$$

† The **intermediate** one in which the "fundamentals" are not so bad but no so good either. In this case a coordination game is created and, consequently, no strategy dominates all others: which one is optimal depends on what other people do. In the experiment, this corresponds to the scenario in which the agency is "medium" ($q = B$): if the other person in my group evades, it is optimal for me to evade as well; if the other person complies, I am better off if I comply too.[10] Formally,

$$x(Y, Y, B) > x(Z) > x(Y, Z, B) \quad (8)$$

Turning now to the other dimension that defines treatments, the difference between the Evasion and Compliance ones is due to their different predictions regarding what a participant's optimal strategy should be. Thus, distinguishing E from C treatments demands the solving of the taxpayer problem, namely, that of choosing between Evasion (Y) and Compliance (Z) using all the information available (s) in order to maximise expected utility. In this setup, therefore, a taxpayer's strategy $\boldsymbol{\sigma}$ is a vector of decisions, one for each possible signal $s \in \mathcal{S}$. Formally, $\boldsymbol{\sigma} := (\sigma(a), \sigma(b), \sigma(c))$, where $\sigma : \mathcal{S} \to \mathcal{D}$ is a function that maps signals into decisions.[11] Therefore, finding the solution requires comparing the (certain) utility of

[10] Clearly, this condition does not apply to the Lottery case.

[11] Actually, it maps signals into *probability distributions* over decisions, if one allows for mixed strategies. However, this possibility was explicitly ruled out here because its inclusion would not have provided any extra, significant insight as to justify the complexity-associated problems it would have entailed.

compliance, $u(Z)$, and the expected utility from evasion,

$$Eu\left(Y,\mathbf{k}'\left(\mathbf{s}'\right)|s\right) := \sum_{q\in\mathcal{Q}} f\left(q|s\right) \sum_{s'\in\mathcal{S}} \Pr\left(s'|q\right)\left\{k'\left(s'\right)u\left(Y,Y,q\right)+\left[1-k'\left(s'\right)\right]u\left(Y,Z,q\right)\right\} \quad (9)$$

where $u(Y,d',q) := u(x(Y,d',q))$ is the utility from receiving payoff $x(Y,d',q)$; $s' \in \mathcal{S}$ and $d' \in \mathcal{D}$ are respectively the signal and decision of the other member of the group; $\Pr(s'|q) \in [0,1]$ is the conditional probability of the other member getting a signal s' given that the agency's type is q; $\mathbf{k}'(\mathbf{s}') := (k'(a), k'(b), k'(c))$; and $k'(s')$ is the optimising person's belief about the probability that the other member of the group will evade given that the latter receives signal s'.

This comparison depends crucially on the beliefs of the optimiser with respect to the actions of the other member of the group, $\mathbf{k}'(\mathbf{s}')$, and, therefore, on the ability and sophistication of subjects at forming them, a matter that is directly related to the concepts of common knowledge and higher-order beliefs (HOBs, Carlsson and van Damme, 1993)). These HOBs refer to the levels of reasoning involved in reaching a conclusion and are neatly connected to the method of Iterated Deletion of Strictly Dominated Strategies (IDSDS): for each iteration, the order of beliefs increases one level. Furthermore, HOBs are the key factor behind the uniqueness of the global game equilibrium: In the first iteration, $i = 1$, my private signal gives me information about the set of strategies (out of the original set, Σ^0) that are strictly dominated (*SDed*) by others and will therefore *never* be played. In the second iteration, $i = 2$, the set of those strategies that survived the previous round of deletions is the new feasible set, Σ^1. Via an analogous mechanism, a new group of *SDed* strategies will be discarded and after that a new iteration $i = 3$ with feasible set Σ^2 will begin. The theory of global games proves that in the limit, after an arbitrarily large number of iterations, the feasible set Σ^∞ has only one element, $\boldsymbol{\sigma}^*$. In other words, the equilibrium is unique.

In the present experiment, only 2 iterations are needed to find the unique solution to the taxpayer problem.[12] Depending on the number of iterations used (1 or 2), a taxpayer is then classified as "Rudimentary" or "Advanced", respectively. Their behaviour is summarised in the following two propositions.

Proposition 1 ***Rudimentary Dominance (RDom)****: According to Rudimentary taxpayers (RTPs): 1. if $s = a$ (signal is low), Evasion strictly dominates (SDs) Compliance: $Y \succ_R Z$; 2. if $s = b$ (signal is medium), no strategy SDs the other: $Y \nsucc_R Z$ and $Z \nsucc_R Y$; and 3. if $s = c$ (signal is high), Compliance SDs Evasion: $Z \succ_R Y$.*

Proposition 2 ***Advanced Dominance (ADom)****: According to Advanced taxpayers (ATPs): 1. those strategies that are rudimentary-dominated (parts 1 and 3 of proposition 1) are also advanced-dominated; and 2. if $s = b$ (signal is medium), then: 2.a. in E treatments,*

[12] This does not apply to Lottery treatments for the obvious reason that in those cases, by definition, a taxpayer's payoff does not depend on other people's choices or the taxpayer's beliefs about them.

Evasion SDs Compliance: $Y \succ_A Z$*; and 2.b. in C treatments, Compliance SDs Evasion:* $Z \succ_A Y$*.*

At this point, it is worth defining the concepts of Soft, Medium and Tough games, which are simply the games played by the members of a group when the agency is soft, medium and tough, respectively (i.e., they are like the game shown in table 2, with $q = A$, B and C). Clearly, these games $g \in \mathcal{G} := \{S, M, T\}$ depend on the type of the agency, and so both g and q are subject to the same probabilistic process.

Based on this taxonomy of games and on the conditional probability distribution of q (shown in table 3), two different scenarios can be identified: one in which the signals give perfect information about the game being played, and another one in which precision is less than perfect.

In the first iteration, therefore, a taxpayer who receives a soft signal ($s = a$) knows for sure that she is playing the Soft game ($g = S$). Furthermore, because of equation 6, she can immediately realise that Evasion *SDs* Compliance, the very result indicated in part 1 of proposition 1. Following a similar argument and using equation 7, part 3 is also proved.

When the signal is medium ($s = b$), though, the person does not know the actual game g that is played, but she does know its conditional probability distribution $f(g(q)|b) = f(q|b)$. Thus, the game that she faces is depicted in figure 1, and her expected utility from evasion is given by equation 9, where s is replaced by b. This expression is an increasing function of the beliefs about the other member's probability of evasion, $k'(s')$, $\forall s' \in \mathcal{S}$, because of the nature of the contingent policy (equation 5). The worst-case scenario for the optimising person occurs, therefore, when she expects the other member to choose Compliance irrespective of the signal received ($\mathbf{k}'(\mathbf{s}') = \mathbf{0}$, $\mathbf{0} := (0, 0, 0)$), such that the expected utility from evasion is $Eu\,(Y, \mathbf{0}|b)$. Analogously, the best case scenario corresponds to that in which the other member always evades ($\mathbf{k}'(\mathbf{s}') = \mathbf{1}$, $\mathbf{1} := (1, 1, 1)$) and expected utility is $Eu\,(Y, \mathbf{1}|b)$. It is not difficult to see that the no-strict-dominance condition of proposition 1 (part 2) requires

$$Eu\,(Y, \mathbf{0}|b) < u(Z) < Eu\,(Y, \mathbf{1}|b) \tag{10}$$

and if it is satisfied, a Rudimentary taxpayer will act exactly as predicted by the *RD* proposition.[13]

A Rudimentary taxpayer would stop her analysis here, but the Advanced one will continue

[13] An alternative interpretation of this equation that will be used later is the following. Let us construct a new, artificial 2x2 game like the one in table 2, but which is a weighted average of the Soft, Medium and Tough games defined above, $A := \sum_{q \in \mathcal{Q}} f\,(q|b) \cdot g\,(q)$, so that the corresponding (expected) utility in each of its cells is

$$u\left(d, d', E\,(q|b)\right) := \sum_{q \in \mathcal{Q}} f\,(q|b) \cdot u\left(d, d', q\right) \tag{11}$$

It can then be shown that $u\,(Y, Z, E\,(q|b)) = Eu(Y, \mathbf{0}|b)$, $u\,(Y, Y, E\,(q|b)) = Eu(Y, \mathbf{1}|b)$, and $u\,(Z, Y, E\,(q|b)) = u\,(Z, Z, E\,(q|b)) = u(Z)$, so that equation 10 implies that this "Average game" is a coordination game.

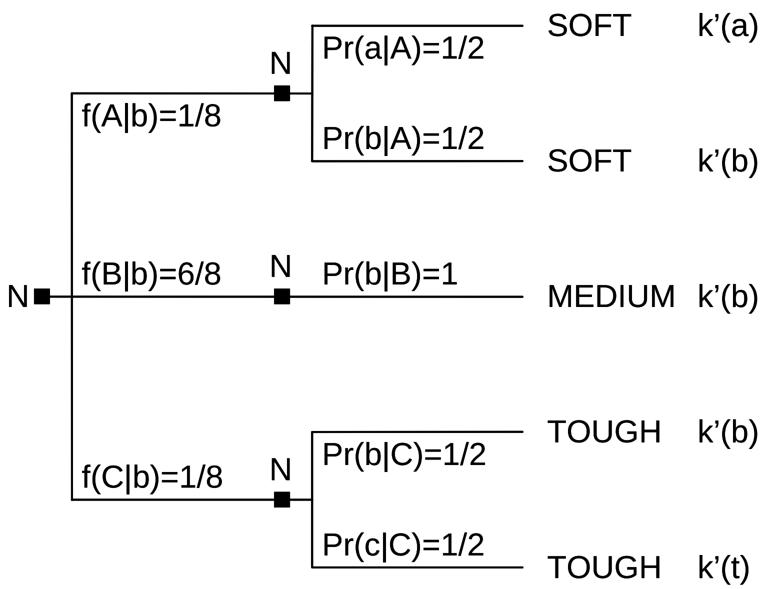


Figure 1: Game tree if signal is medium ($s = b$).

to the next iteration. Furthermore, the ATP will realise that, if the other member of her group is (at least) Rudimentary, then (by symmetry) she would have also worked out that Evasion (respectively, Compliance) is the strictly dominant strategy when the signal received is soft (a) (respectively, tough (c)). Formally, the ATP's beliefs about the other person's choices will have precise numbers attached to them, namely, $k'(a) = 1$ and $k'(c) = 0$. The expected utility will reflect this in general, $Eu\left(Y, \left(1, k'\left(b\right), 0\right)|b\right)$, as well as in the worst- and best-case scenarios, $Eu\left(Y, \mathbf{c}|b\right)$ and $Eu\left(Y, \mathbf{e}|b\right)$,where $\mathbf{c} := (1, 0, 0)$ and $\mathbf{e} := (1, 1, 0)$.

Depending on the position of the safe utility $u(Z)$ with respect to the latter two, three cases can arise, of which we are interested only in the following two:[14]

$$Eu\left(Y, \mathbf{c}|b\right) > u(Z) \tag{12}$$
$$u(Z) > Eu\left(Y, \mathbf{e}|b\right) \tag{13}$$

The first one indicates that even in the worst-case scenario, the expected utility from Evasion is higher than that of Compliance or, equivalently, that the former SDs the latter. The second one, on the other hand, implies that, even in the best-case scenario, the expected utility from Evasion is lower than that of Compliance, and so that the second SDs the first.

By definition these two conditions are mutually exclusive, and which one of them is satisfied

[14]The third one does not lead to a unique solution, which goes against the spirit of the theory of global games. The reason for the non-uniqueness is the discreteness of the model. Having continuous choices may have avoided this problem, but at the cost (considered to be too high) of increasing the complexity of the game and thus the noise in the observations.

determines the taxpayer's optimal strategy: either $\boldsymbol{\sigma}^* = (Y, Y, Z)$ if equation 12 holds or $\boldsymbol{\sigma}^* = (Y, Z, Z)$ if the one that holds is number 13. These strategies are of the "threshold" type (Heinemann et al (2004a), Heinemann et al (2004b)) but can be indexed by their second component, which is the only one that differentiates one strategy from the other and corresponds to the optimal choice when the signal is medium, $\sigma^*(b)$. The value of this component, therefore, is the one that defines the Evasion, $\sigma^*(b) = Y$, and Compliance, $\sigma^*(b) = Z$, treatments.

The rationale for including these two types of treatments reflects, above all, the lack of theoretical predictions or stylised facts about what strategy we should expect to be played in the medium case. Its presence, however, allows for the testing of some hypotheses regarding the comparative statics of global games: in particular, the change of parameters predicts that the number of people receiving b signals that choose Y should be greater in E-treatments than in C-ones, while no significant difference should exist if signals are soft (a) or tough (c).[15] This is summarised as follows:

Hypothesis 1 ***Optimal strategy (OS)****: 1. If signal is soft ($s = a$) then evade ($d = Y$); 2. if signal is tough ($s = c$) then comply ($d = Z$); and 3. if signal is medium ($s = b$) then: 3.a. in E treatments, evade ($d = Y$); and 3.b. in C treatments, comply ($d = Z$).*

If choices satisfy all three parts of the hypothesis, then one can say they are "consistent with the *ADom* predictions" and label the taxpayer as "Advanced". If they only satisfy the first two parts, they are "consistent with the *RDom* predictions" and the taxpayer can be labelled as "Rudimentary".

3.2 Selection of payoffs

Turning now to the main prediction of the TEGG model, it is clear that, in order to test which of the auditing rules (contingent or cut-off) is the best of the two, one needs a "level playing field". In this setup, it requires the enforcement costs to be the same in G- and L-treatments, which further simplifies to undertaking the same (expected) number of audits (for a given value of q) in each treatment. This way, the mere comparison of the errors made by each type of agency across treatments will indicate which rule is superior to the other (if any). Formally,

[15] For L-treatments, the analysis is greatly simplified since other people's choices do not affect one's decisions. Then, the equivalents of equations 12 and 13 are, respectively,

$$Eu(Y, b) > u(Z) \quad (14)$$

$$u(Z) > Eu(Y, b) \quad (15)$$

Hypothesis 2 ***Superiority of Contingent Rule (SCR):*** *Given a fixed level of enforcement, Global treatments yield less (expected) errors than Lottery ones for all possible types of agency,* $q \in \mathcal{Q}$.

The expected number of audits is

$$Ea\left(d, d', q\right) = \sum_{y \in \mathcal{Y}} \Pr\left(y\right) \sum_{s \in \mathcal{S}} \Pr\left(s|q\right) \sum_{s' \in \mathcal{S}} \Pr\left(s'|q\right) \sum_{d \in \mathcal{D}} \Pr\left(d|s, y\right) \sum_{d' \in \mathcal{D}} \Pr\left(d'|s', y\right) \sum_{a \in \mathcal{A}} \Pr\left(a|d, d', q\right) \sum_{a' \in \mathcal{A}} \Pr\left(a'|d, d', q\right) \cdot \left(a + a'\right) \quad (16)$$

where y is the income of the population, that can be 0 ("bad" year) or 1 ("good" year), $\mathcal{Y} := \{0, 1\}$ and $\mathcal{A} := \{0, 1\}$.[16]

Thus, the equalisation of enforcement costs demands that, for each and every value of the agency's type $q \in \mathcal{Q}$, the expected number of audits in G-treatments must be equal to the corresponding one in L-treatments. Since payoffs are linear functions of the probabilities, the equalisation requires the following conditions to hold:

LE	GE	LC	GC
$u\left(Y, A\right) =$	$u\left(Y, Y, A\right)$	$u\left(Y, A\right) =$	$\sum_{d' \in \mathcal{D}} \omega\left(d'\right) \cdot u\left(Y, d', A\right)$
$u\left(Y, B\right) =$	$u\left(Y, Y, B\right)$	$u\left(Y, B\right) =$	$u\left(Y, Y, B\right)$
$u\left(Y, C\right) =$	$\sum_{d' \in \mathcal{D}} \omega\left(d'\right) \cdot u\left(Y, d', C\right)$	$u\left(Y, B\right) =$	$u\left(Y, Y, C\right)$

Note: The weights, w(d'), d'∈D, are functions of the parameters of the problem. In the experiment, w(Y)=1/6 and w(Z)=5/6.

Table 5: Payoffs. Conditions for the equalisation of enforcement costs.

Focusing on the Evasion case (columns 1 and 2), it requires that the payoff of evaders when the agency is soft (A) or medium (B) should be the same regardless of the auditing rule. This is logical because in these cases taxpayers can only get soft or medium signals (a or b) and, because of Rudimentary and Advanced dominance (equations 6 and 12/14), they will always evade. Since the agency cannot tell whether the pair of low declarations is the result of a bad year or of evasion, it will audit with the same probability in both G and L treatments (no difference in the information available to the agency) and so the (expected) payoffs of taxpayers are the same as well. When the agency is tough, however, taxpayers could receive medium or tough signals, $s \in \{b, c\}$, with those receiving $s = b$ choosing Y (because of equation 12/14) and the others choosing Z (equation 7). In the L-treatments the agency only uses the information derived from an individual's action; in G-treatments, on the other hand, it also uses the one derived from the other person's choice. In particular,

[16] In the experiment, however, implicit in the stage game (table 2) is the assumption that all taxpayers have high income ($y = 1$). The reasons for this are that introducing the possibility of low income periods will not add to our knowledge (trivially, if $y = 0$ everyone declares truthfully) and that all interesting hypotheses to test are related to the high-income scenario (not to mention the extra cost and time that running this expanded experiment will demand).

if choices are different from each other, the agency can be sure that the one who chose Y is likely to be lying and can therefore audit her with a higher probability and the other one with a lower one. This is the basic mechanism behind the contingent rule and relies heavily on the extra difficulty taxpayers face when trying to coordinate on the Full Evasion equilibrium (strategic uncertainty). The actual position of $u(Y, C)$ between $u(Y, Y, C)$ and $u(Y, Z, C)$ is determined by the weights $\omega(d') \in [0, 1]$, $d' \in \mathcal{D}$, which depend on the parameters of the problem (especially $p(q)$ and $f(q|s)$) and in the experiment take the values 1/6 (for $d' = Y$) and 5/6 (for $d' = Z$). The conditions for the C-treatments (columns 3 and 4) are found following a similar argument.

The parameters chosen for the four treatments are therefore the ones shown in table 6.

Person 1's choice	Person 2's choice	Type of agency	GC	GE	LC	LE
Y	Y	A	1,000	1,000	715	1,000
Y	Y	B	655	145	655	145
Y	Y	C	579	6	579	1
Y	Z	A	658	156	715	1,000
Y	Z	B	651	135	655	145
Y	Z	B	0	0	579	1
Z	{Y,Z}	{A,B,C}	654	140	654	140

Note: Only payoffs of Person 1 are shown. Those of Person 2 are symmetric.

Table 6: Payoffs. All treatments.

These payoffs satisfy all the conditions mentioned so far: the "type of agency" one (equation 3), the "global game" ones (equations 6, 7 and 8), the "average game" one (equation 10), the "equal enforcement costs" ones (table 5), and the ones that define treatments: $L\ v\ G$ (equations 4 and 5) and $E\ v\ C$ (equations 12/14 and 13/15).

The actual vector of values chosen is just one among many that satisfy the above mentioned conditions. The feasible set was narrowed down by setting, without loss of generality, the maximum and minimum payoffs in the G-treatments equal to $1,000$ and 0 respectively, and by restricting attention to natural numbers.[17] Noting that payoffs in L-treatments are deterministic functions of those in G-treatments (see table 5), only 10 parameters remain to be determined, namely, the intermediate payoffs of GC and GE (including the safe payoffs). Before getting to it, however, a digression about equilibrium selection is in order here.

The global game (GG) technique selects one of the equilibria of a coordination game, an equilibrium that coincides (for 2×2 games like the ones used here) with the one selected by the "risk dominance" criterion (RD, Harsanyi and Selten, 1988)). Intuitively, the latter chooses the equilibrium which, if abandoned, inflicts the highest costs on the players. Since the criterion applies to the global game we need to consider all 3 possible scenarios it usually entails: the 2 extreme ones and the intermediate one mentioned in the previous section. In

[17] To simplify computations and understanding by subjects, as well as to avoid prospect-theoretical interpretations (which, though interesting in themselves, are not the focus of the present analysis).

this particular case, however, it is enough to concentrate on the "average game" (defined in footnote 13), since it neatly summarises the whole game and thus simplifies the analysis. Because this "average game" is a coordination game, it will have 2 pure-strategy equilibria: one in which both players choose Y and get $Eu(Y, \mathbf{1}|b)$, and another one in which they both choose Z and get $u(Z)$. Which of the two is the risk-dominant one depends on the relationship between $l_E := Eu(Y, \mathbf{1}|b) - u(Z)$ (the loss from deviating from the Full Evasion equilibrium) and $l_C := u(Z) - Eu(Y, \mathbf{0}|b)$ (the loss from deviating from the Full Compliance one). If deviating from (Y, Y) is more costly than deviating from (Z, Z) (i.e., if $l_E > l_C$), the risk-dominant equilibrium (RDE) is (Y, Y); otherwise, it is (Z, Z). In the experiment, the RDE depends on the treatment: it is (Y, Y) in GE and (Z, Z) in GC. These are, not surprisingly, the choices that equations 12 and 13 predicted to be optimal in those treatments, thus confirming that both the GG theory and the RD criterion select the same equilibrium.

There is, however, an important competitor for the RD/GG criterion: the payoff-dominance one (PD). It simply states that if all equilibria can be Pareto-ranked, players will coordinate on the dominant one. In the experiment, the payoff-dominant equilibrium (PDE) is always (Y, Y) regardless of the treatment, because of the Average game being a coordination game and the contingent policy penalising evaders in case of coordination failure.

Thus, the PD and RD criteria select the same equilibrium (Y, Y) in the GE treatment but different ones ((Y, Y) and (Z, Z), respectively) in the GC one. Since the criteria reinforce each other in GE but compete against each other in GC, this suggests an interesting hypothesis to test:

Hypothesis 3 ***Relative frequency GE/GC (RF)****: The frequency of choices consistent with the GG/RD prediction will be (weakly) higher in GE than in GC.*

The main hypothesis of interest, however, is whether data fits the GG predictions (hypothesis 1). Thus, the 10 "free" parameters in table 6 were chosen to make the satisfaction of the predictions as difficult as possible, i.e., by making the RDE as little risk-dominant as possible. This required minimising l_E and maximising l_C in GE,and the opposite in GC. This way, if the data supports the GG predictions in these most demanding conditions, then the theory could be expected to be an even better predictor in more favourable environments.

Finally, it is important to mention here that risk aversion could dramatically alter the predictions of the model, and this may be especially important since evidence indicates that attempts to induce risk-preferences seem not to work (Selten *et al.*, 1999)). The solution implemented in the experiment was to choose parameters such that all constraints will be satisfied for a large range of risk preferences. In particular, in E-treatments the parameters of table 5 are robust to degrees of relative risk aversion as high as 0.4 (about 60% of the population, according to Holt and Laury, 2002)). For C-treatments, they are robust for values as low as 0 (about 80% of the population, according to the same study). Also, it is acknowledged in the experimental literature that when playing complex games people often avoids the complications of utility maximisation and instead simply maximise payoffs, which implies that risk preferences should not be an important issue here (probably most of the participants will end up acting as if their degrees of risk aversion were somewhere in the $[0, 0.4]$ range).

4 Results

A total of 1,520 observations were collected in the experiment, and table 7 shows the breakdown by treatment. It also shows summary statistics of the key variables needed for testing the hypotheses of the previous section: "Dominance" and "Errors". The first one measures the coincidence between the data and GG theoretical predictions about the subjects' choices (DOM=1 if data fits predictions and 0 otherwise). Its name reflects the fact that those predictions are based on the concepts of dominance (propositions 1 and 2). The second one quantifies the number of errors (per observation/datapoint) made by the agency (ERR=1 if an error was made, 0 otherwise). Note that Dominance is never lower than 50% and Errors never above 35%.

Treatment	Observations	Dominance (DOM)		Errors (ERR)	
		Mean	St. Dev.	Mean	St. Dev.
GC	360	0.7722	0.4200	0.1522	0.2252
GE	360	0.8639	0.3434	0.2028	0.3034
LC	400	0.5450	0.4986	0.3473	0.3303
LE	400	0.9300	0.2555	0.3243	0.3726
All	1,520	0.7757	0.4173	0.2608	0.3248

Note: DOM=1 if subject's choice coincides with GG's prediction, 0 otherwise. Error=1 if agency made an error, 0 otherwise.

Table 7: Summary Statistics. Dominance and Errors.

For hypothesis testing, it would be useful to aggregate data in two different ways, depending on the information available to the relevant actor. Thus, for hypotheses related to the decisions of the taxpayers (OS and RF), data are aggregated by signal (columns 3-5 in table 8). For those related to actions of the agency (SCR), on the other hand, the aggregation is done according to the type of agency (columns 6-8 in the same table).

Variable	Role	Type	Description
DOM	Dependent	Dummy	1 if choice coincides with prediction, 0 otherwise
DOMs	Dependent	Dummy	Idem DOM, but for a fixed $s \in \mathcal{S}$
RD	Dependent	Dummy	Idem DOM, but for $s \in \{a, c\}$
AD	Dependent	Dummy	Idem DOM, but for $s = b$
g	Explanatory	Dummy	1 if G treatment, 0 otherwise
e	Explanatory	Dummy	1 if E treatment, 0 otherwise
ge	Explanatory	Dummy	Interaction term: 1 if GE treatment, 0 otherwise
a	Explanatory	Dummy	1 if $s = a$, 0 otherwise
b	Explanatory	Dummy	1 if $s = b$, 0 otherwise
c	Explanatory	Dummy	1 if $s = c$, 0 otherwise

Note: "Predictions" as defined in hypothesis 1.

Table 9: Variables of the model. Dominance.

Treatment	Observations	Signal (s)			Agency's type (q)		
		a	b	c	A	B	C
GC	360	7	295	58	18	234	108
GE	360	29	292	39	54	234	72
LC	400	29	330	41	60	260	80
LE	400	51	337	12	100	280	20
All	1,520	116	1,254	150	232	1,008	280

Note: Interpretation of s/q: a/A: "soft"; b/B: "medium"; c/C: "tough".

Table 8: Number of observations, aggregated by signal and type of agency.

For the analysis, data from all subjects for all periods were pooled. This is justified by the fact that there is little variability in behaviour after the first few rounds of each treatment,[18] with many people choosing exactly the same option every time they receive a given signal. This lack of variability over time is not a bad thing in itself (since the theory actually predicts such rigidity), but it precludes the possibility of using other econometric techniques (e.g., panel data).

4.1 OS and RF hypotheses

The set of variables that is going to be used for testing is described in table 9.

[18]Except in the GE one, that requires 10 rounds to become stable. This, however, does not usually have an impact on results, and when it does, it will be mentioned in the text.

Dep. Var.:	DOMa	DOMb	DOMc	RDOM	ADOM	DOM
a				1.0205		0.7091
				[0]		[0]
b						0.5030
						[0]
c				0.9855		0.7671
				[0]		[0]
g	0.0000	0.2803	-0.0345	-0.0201	0.2803	0.2227
	[0.082]	[0]	[0.158]	[0.35]	[0]	[0]
e	-0.0196	0.4714	0.0000	-0.0297	0.4714	0.3928
	[0.323]	[0]	[0.706]	[0.072]	[0]	[0]
ge	0.0196	-0.3543	-0.0681	-0.0095	-0.3543	-0.2998
	[0.323]	[0]	[0.217]	[0.804]	[0]	[0]
cons	1.0000	0.4485	1.0000		0.4485	
	[.]	[0]	[0]		[0]	
Obs	116	1,254	150	266	1,254	1,520
LC	1.0000	0.4485	1.0000	1.0000	0.4485	.5450
LE	0.9804	0.9199	1.0000	0.9841	0.9199	.9300
GC	1.0000	0.7288	0.9655	0.9692	0.7288	.7722
GE	1.0000	0.8459	0.8974	0.9412	0.8459	.8639

Note: Top panel: Probability that estimate =0 is shown in brackets below estimate. Bottom panel displays estimated average values of the dependent variable for each treatment.

Table 10: Estimation. Dominance. Overall and by signal.

DOMs measures Dominance when only observations with a given signal s are considered. RDOM means Rudimentary Dominance and considers only observations when signals are soft (a) or tough (c). ADOM measures Advanced Dominance and only takes into account observations with medium signals (hence, it is identical to DOMb). The model used for testing is then

$$DOM = \alpha + \beta_1 g + \beta_2 e + \beta_3 ge + \gamma_1 a + \gamma_2 b + \gamma_3 c + \varepsilon \tag{17}$$

(analogous ones are used for the alternative dependent variables) and the estimates are shown in table 10.

Table 11 shows the results of the test in a schematic way.[19] The first panel tests the OS hypothesis (see note below the table for interpretation of symbols). The null hypothesis is that data are consistent with GG predictions,[20] a hypothesis that is supported in the cases of low and high signals ($s = a$ or c) and that implies that people are, at least, Rudimentary.[21] When the signal is medium, however, the GG's predictions are rejected for all treatments and, therefore, the OS hypothesis is quantitatively rejected as well (i.e., those aspects related

[19] The tests are shown in table 18 in appendix B.

[20] The predicted value (following proposition 2, AD) is 1 for all cases, which is interpreted as requiring that all observations should match predictions.

[21] The null hypothesis is rejected in the GE case because of an outlier. If eliminated, the hypothesis cannot be rejected.

to part 2 of the hypothesis). Qualitatively, however, the results do support the predictions, as can be seen in figures 2 and 3, where the observed strategies resemble the shape of the predicted ones (except for LC, that will be analysed in detail later).[22] Having in mind the discreteness of the model (which amplifies divergences) and that the parameters were chosen to make the test as difficult to pass as possible for the GG theory, the result is still encouraging.

Dep. Var.:	DOMa	DOMb	DOMc	RDOM	ADOM	DOM
LC		X			X	X
LE		X			X	X
GC		X			X	X
GE		X	X	X	X	X
LC=GC		GC			GC	GC
LE=GE	GE	LE	LE		LE	LE
LC=LE	LC	LE			LE	LE
GC=GE		GE			GE	GE

Note: Top panel: Empty if data fits prediction in hypothesis OS; "X" otherwise. Bottom panel: Empty if no difference, treatment with higher dominance otherwise.

Table 11: Dominance tests. Predictions and inter-treatment comparisons.

Result 1 *Qualitative ADom (QAD): People are, at least, Rudimentary: they act as predicted by the RDom proposition when signals are low or high. The hypothesis that they make decisions in a way consistent with the ADom predictions is quantitatively rejected (and so is the OS hypothesis, consquently) but supported qualitatively.*

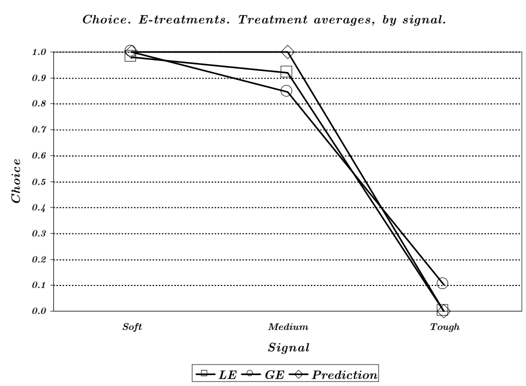


Figure 2: Observed and Predicted choices. E-treatments.

The bottom panel of table 11 compares the levels of Dominance across treatments. The null hypothesis for the first two lines is that Dominance is the same in Global and Lottery treatments, a hypothesis that (following the general result) is supported for $RDom$ but not for $ADom$. The result that the G/L comparison depends on whether E or C is played, on

[22] In the figures, 1 corresponds to Evasion (choice Y) and 0 to Compliance (choice Z).

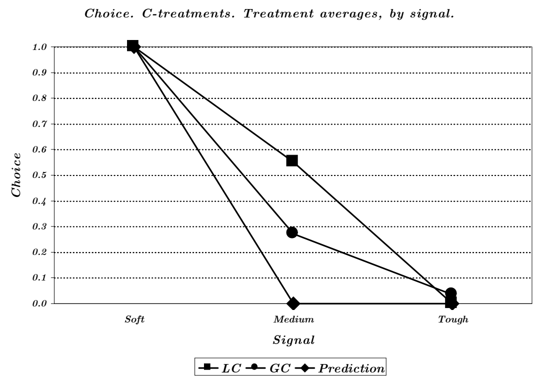


Figure 3: Observed and Predicted choices. C-treatments.

the other hand, is something that the theory cannot explain (there should be no difference, theoretically). It is important to mention, though, that the difference between GE and LE is drastically reduced if considering only the last 10 periods of the session (see figure 5), so it could be said that Global treatments foster more Dominance than Lottery ones.

For the last two lines, the null hypothesis is that Dominance is the same in Evasion and Compliance treatments. Once again, $RDom$ is satisfied but $ADom$ is not, but in the latter case the results are clear now: E treatments are more consistent with predictions than C ones. This can be explained by the coincidence of RDE and PDE in the former ones and the discrepancy between them in the latter ones.[23] This is therefore consistent with the RF hypothesis.

Result 2 *Comparison RDE-PDE (R/PDE): Choices in E treatments are consistent with predictions more frequently than in C ones.*

These results can also be visualised in figures 4 and 5. The first one confirms that $RDom$ is strongly supported by data and that different treatments do not affect it. The second one focuses on choices when the signal is medium and attests that $ADom$ predictions are quantitatively rejected, though qualitatively supported. It also shows that treatments can be ranked as determined by the tests, namely, (from higher to lower Dominance), LE, GE, GC and LC.[24]

[23] Actually, this only applies to the Global treatments, since clearly there is no coordination game in Lottery ones. There is no similar explanation for the difference between E and C in the latter ones.

[24] Restricting attention to the last 10 periods so that the learning process in GE converges, the difference between GE and LE vanishes. This is consistent with the RDE/PDE argument, since it seems that people learn to play the only "reasonable" equilibrium.

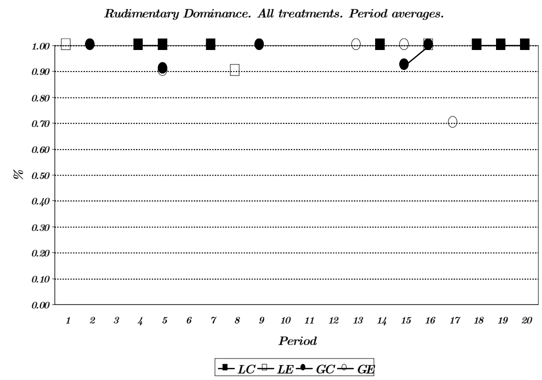


Figure 4: Rudimentary Dominance. All treatments. Period averages

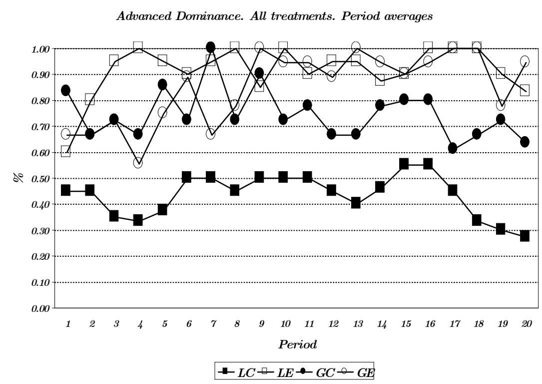


Figure 5: Advanced Dominance. All treatments. Period averages.

4.2 Characteristics and Decisions

The analysis of choices can be furthered by using the information collected in the questionnaire run after the experimental rounds. The relevant variables are shown in table 12.

Variable	Role	Type	Description
ADOM	Dependent	Dummy	1 if data fits prop. AD (part 2), 0 otherwise
g	Explanatory	Dummy	1 if G treatment, 0 otherwise
e	Explanatory	Dummy	1 if E treatment, 0 otherwise
ge	Explanatory	Dummy	Interaction term: 1 if GE treatment, 0 otherwise
gender	Explanatory	Dummy	1 if female, 0 otherwise
age	Explanatory	Natural	
study	Explanatory	Dummy	0 : no study, 1 : non-economics, 0 : economics
# exp	Explanatory	Dummy	0 : none, 1 : 1 to 4, 2 : 5+ experiments
math	Explanatory	Dummy	0 : none, 1 : basic, 2 : advanced knowledge
prob	Explanatory	Dummy	0 : none, 1 : basic, 2 : advanced knowledge
game	Explanatory	Dummy	0 : none, 1 : basic, 2 : advanced knowledge

Note: "Study" refers to "area of study". "Math"/"Prob"/"Game" refer to knowledge of mathematics, probability theory and game theory, respectively.

Table 12: Questionnaire variables. Dominance.

The analysis will be restricted to that of $ADOM$. The reasons for this are two: first, the previous section proved that RD is satisfied almost perfectly for the whole sample of participants, regardless of their individual characteristics; and second, $ADOM$ is the main source of DOM variability, since in most observations the signal is medium (see table 8).

The questionnaire also asked participants about the strategies they followed and the rationale behind them. This information was then used to classify them according to some stylised characteristics, in a fashion similar to the one used by Bosch-Domenech *et al.*, 2002). The distribution of subjects in terms of categories and treatments is shown in table 13.

Category	GC	GE	LC	LE	All
Expected payoff maximisers (EPM)	10/11	8/11	5	5/13	28/40
Chance maximisers (CM)	1/2	0/3	6/7	0/8	7/20
Learners (L)	0	3	1	1	5
Mixers/Experimenters (M/E)	1	2	0	2	5
Non-independent (NI)	1	0	4	3	8
Randomisers (R)	1	2	1	0	4
Confused (C)	1	0	1/2	1	3/4
Risk-lovers (RL)	2	0	1	0	3
All	18	18	20	20	76

Note: Cells with two numbers separated by "/" reflect uncertainty about the allocation of some subjects to specific categories.

Table 13: Questionnaire. Classification of subjects.

The different categories are defined as follows:[25]

- Expected payoff maximisers (EPM): Those who indicated they played either Y in E treatments or Z in C ones, based on expected-payoff maximisation. Note that this

[25] Appendix C shows comments from some subjects' questionnaires that are characteristic of each one of these categories.

categories includes everyone who played according to the OS strategy, even though they did not use HOBs.

- Chance maximisers (CM): Those who only considered the probabilities of outcomes being higher or lower than the safe option, without weighting them using the associated payoffs.
- Learners (L): Those whose decisions varied in the first periods, but chose always the same action afterwards.
- Mixers/Experimenters (M/E): Those that deviated just once or twice from the predictions of the OS hypothesis but, unlike the Ls, did so at times other than the first periods (Experimenters). An alternative rationale could be that they followed a strategy such that they evaded and complied with probabilities that usually replicated the relevant odds ($(1/8, 7/8)$ in C treatments and $(7/8, 1/8)$ in E ones), and so could be labelled "Mixers".
- Non-independent (NI): Those who (despite the instructions clearly stating that rounds were independent from each other) followed some kind of history-dependent strategy.
- Randomisers (R): Those who chose randomly between Y and Z. Also called "Guessers" (G).
- Confused (C): Those who seemed to be (or acknowledge they were) confused.

On top of these strategies, the degree of risk aversion is expected to play a role as well. In particular, risk aversion fosters compliance (*ceteris paribus*) and hence makes GG's predictions easier to be satisfied in C treatments, but works against them in E ones. Combining the strategies defined above and the degree of risk aversion, one can usually categorise all subjects and find some interesting stylised facts.

The first one that can be stated is that categories seem to order themselves in three "Dominance bands" according to their degree of coincidence with the GG predictions (see figures 6, 7, 8 and 9). Near the top we can find the $EPMs$ (high dominance). In the middle-ground there is a mixed bag of types (M/E, L, NI and C) who chose different actions in different periods, even though they always got the same signal b. Risk lovers (RL) are close to the top in E treatments and to the bottom in C ones, and the opposite is true for risk averse (RA) people.

All these results, however, are not surprising. The category that is really exciting to analyse in detail, on the other hand, is that of the CMs, since it is behind the case the largest deviations from predictions (GC treatment). Now, the first thing to notice is that in some cases CMs cannot be distinguished from $EPMs$, because the observed data are consistent with the predictions of both criteria (expected-payoff and probability maximisation) and the questionnaire information is vague (this is the rationale for the ambiguity in table 13). For

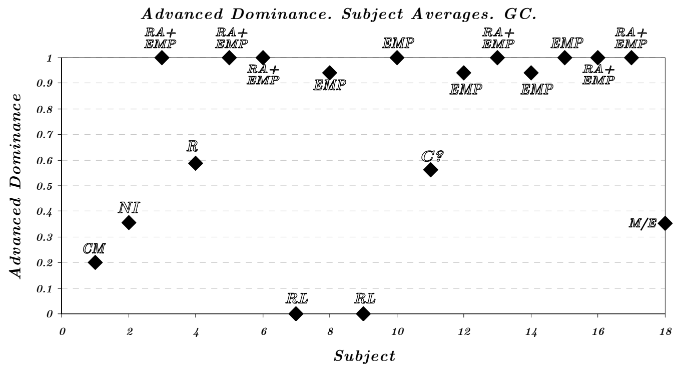


Figure 6: Advanced Dominance. Subject averages. GC treatment.

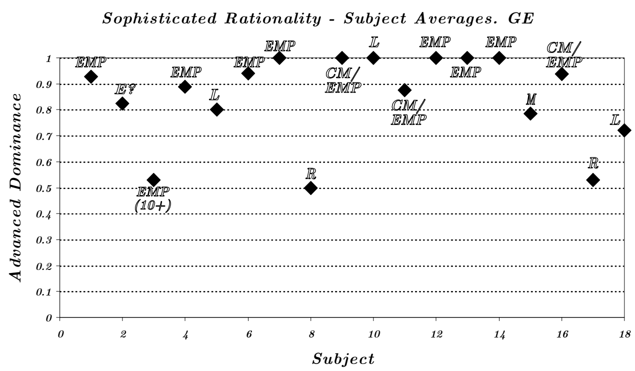


Figure 7: Advanced Dominance. Subject averages. GE treatment.

this very reason, the most interesting scenarios are those where the two criteria prescribe different actions, as is the case in C treatments (GG predicts Compliance, CM predicts Evasion). Focusing on these treatments, it can be seen that significant deviations from the GG predictions take place, thus confirming the results of the tests that compare the levels of dominance in C and E treatments (table 11). Also, since CM's prescription to evade depends on what the other person does in GC but not in LC, it is not surprising that the degree of dominance in the former is greater than in the latter: the uncertainty about the other person's action in GC works against the incentives to evade and (as seen in figure 6) only RLs end up evading all periods. Since this interdependence does not play a role in LC, the number of subjects that evade all periods is far greater (see figure 8), and explains the huge divergence between predictions and data (and confirms the ranking of treatments according to Dominance found in the previous section).

The stylised facts shown so far give us a snapshot of the data, but the question that subsides is: what is behind these choices? what (if any) are the personal characteristics that drive

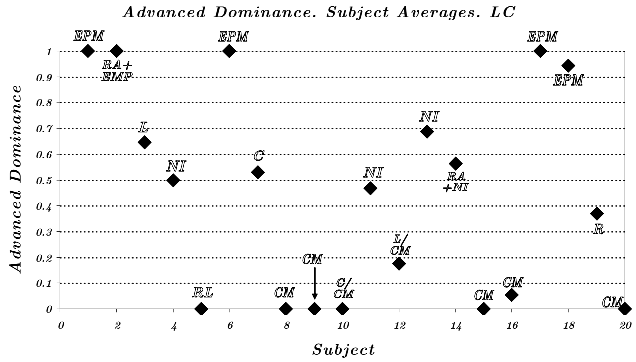


Figure 8: Advanced Dominance. Subject averages. LC treatment.

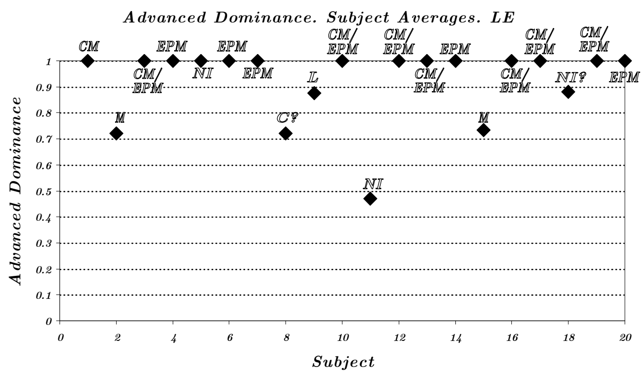


Figure 9: Advanced Dominance. Subject averages. LE treatment.

them? In order to answer these very questions, the variables defined in table 12 were used to estimate the following model:

$$ADOM = \alpha + \beta_1 g + \beta_2 e + \beta_3 ge + \\ + \gamma_1 gender + \gamma_2 age + \gamma_3 \# \exp + \gamma_4 math + \gamma_5 prob + \gamma_6 game + \varepsilon \quad (18)$$

The results (shown in table 14) indicate that estimates are robust to the specification of the model (last three columns)[26] and that most of the times there is not much difference between treatments or between individual treatments and the whole sample. The analysis finds that being male, young, bad-at-maths and good-at-game-theory makes a subject more likely to make decisions that coincide with the *GG* predictions. There is no rationale for the gender effect (which, apart from the whole sample, is only significant in one treatment, anyway), though it is important to note that a similar result is found by Heinemann *et al.*, 2004a). The age effect may seem to reflect that most subjects are university students,

[26] For this very reason, only OLS estimates are shown throughout the whole paper.

but actually it is driven by a few older outliers: if the analysis restricts its attention to "up-to-25-year-olds" (1,050 observations), age becomes non-significant. A similar story can be told about mathematics: it becomes insignificant when the "young" sample is used (thus eliminating the puzzling result that the estimate's sign was negative). Area of study is not significant and, surprisingly, knowledge of probability theory or participation in other experiments are not either (though Heinemann *et al.*, 2004b) find the same result regarding experience).

	OLS					Probit	Logit
	GC	GE	LC	LE	All	All	All
g					0.2914	0.8573	1.3932
					[0]	[0]	[0]
e					0.4895	1.7349	3.0951
					[0]	[0]	[0]
ge					-0.3894	-1.4149	2.4897
					[0]	[0]	[0]
gender	-0.0306	0.0078	0.0967	-0.0738	-0.0616	-0.2761	0.4752
	[0.745]	[0.853]	[0.31]	[0.005]	[0.011]	[0.003]	[0.004]
age	-0.0385	-0.0282	-0.0251	0.0039	-0.0078	-0.0304	0.0540
	[0]	[0]	[0]	[0]	[0]	[0]	[0]
study	0.1916	0.0349	-0.5397	-0.0149	-0.0306	-0.1446	0.2348
	[0.004]	[0.514]	[0]	[0.787]	[0.369]	[0.312]	[0.382]
#exp	0.0006	0.1402	-0.0677	-0.0160	-0.0059	-0.0416	0.0729
	[0.988]	[0]	[0.161]	[0.575]	[0.738]	[0.513]	[0.507]
maths	-0.5119	0.0418	0.1660	-0.0788	-0.0993	-0.3749	0.6982
	[0]	[0.432]	[0.196]	[0.099]	[0.002]	[0.001]	[0.002]
prob	-0.0454	-0.0060	0.0033	0.1204	0.0047	0.0219	0.0314
	[0.635]	[0.893]	[0.963]	[0.007]	[0.866]	[0.834]	[0.867]
game	0.3409	0.1840	0.0362	0.0242	0.0961	0.4249	0.6918
	[0]	[0]	[0.464]	[0.193]	[0]	[0]	[0]
cons	2.0118	1.2868	1.3768	0.8555	0.7772	1.1536	2.1058
	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Obs	295	292	330	337	1,254	1,254	1,254

Note: Probability that estimate =0 is shown in brackets below estimate.

Table 14: Estimation. Effect of personal characteristics on choices.

The only robustly significant variable seems to be knowledge of game theory, which increases Dominance. Furthermore, it is significant in both treatments in which strategic (i.e., game theoretic) interactions took place. This may indicate that some degree of indoctrination may have played a role and so that training can breed "sophistication". This suggests that a typical population (where average knowledge of game theory is negligible) would make choices quite different from the ones suggested by the GG theory. However, if one considers that firms (which are the relevant actors in Sanchez Villalba's model) are sophisticated, then the theory should be a good predictor of behaviour. Moreover, a similar result could be achieved if individual taxpayers had access to sophisticated professional advice, something that is indeed likely to occur.

4.3 SCR hypothesis

The key prediction of the TEGG model is that an agency would be advised to use the contingent auditing rule and to discard the cut-off one. Following Sanchez Villalba, this means that the agency would make less targeting errors if implementing the former than if using the latter, given that enforcement costs are the same in both cases. These errors are the Zeal and Negligence ones defined in section 2 (see especially footnote 2), though –for the reasons explained in footnote 16– the analysis will focus on the latter type only.

The expected loss of a 2-person group can then be expressed as

$$EL(d, d', q) = \sum_{y \in \mathcal{Y}} \Pr(y) \sum_{s \in \mathcal{S}} \Pr(s|q) \sum_{s' \in \mathcal{S}} \Pr(s'|q) \sum_{d \in \mathcal{D}} \Pr(d|s, y) \sum_{d' \in \mathcal{D}} \Pr(d'|s', y)$$
$$\sum_{a \in \mathcal{A}} \Pr(a|d, d', q) \sum_{a' \in \mathcal{A}} \Pr(a'|d, d', q) \left[(1-a)(1-d)y + (1-a')(1-d')y\right] \quad (19)$$

where the expression in square brackets is the sum of negligence errors for the 2-person group. Armed with this information, the model to be estimated is therefore

$$ERR = \beta_1 g + \beta_2 e + \beta_3 ge + \gamma_1 A + \gamma_2 B + \gamma_3 C + \varepsilon \quad (20)$$

where the variables are defined as in table 15.

Variable	Role	Type	Description
ERR	Dependent	Dummy	1 if an error was made, 0 otherwise
ERRq	Dependent	Dummy	Idem ERR, but for a fixed $q \in \mathcal{Q}$
g	Explanatory	Dummy	1 if G treatment, 0 otherwise
e	Explanatory	Dummy	1 if E treatment, 0 otherwise
ge	Explanatory	Dummy	Interaction term: 1 if GE treatment, 0 otherwise
A	Explanatory	Dummy	1 if $q = A$, 0 otherwise
B	Explanatory	Dummy	1 if $q = B$, 0 otherwise
C	Explanatory	Dummy	1 if $q = C$, 0 otherwise

Note: ERR measures errors per person in a 2-person group.

Table 15: Variables of the model. Errors.

The estimates can be seen in table 16.

Dep. Var.:	ERRA	ERRB	ERRC	ERR
A				0.8610
				[0]
B				0.2886
				[0]
C				0.1526
				[0]
g	0.0059	-0.1610	-0.1847	-0.1242
	[0.956]	[0]	[0]	[0]
e	0.3718	-0.2130	-0.1950	-0.1007
	[0]	[0]	[0]	[0]
ge	-0.0967	0.1466	0.1856	0.0804
	[0.423]	[0]	[0]	[0]
cons	0.5482	0.3477	0.1954	
	[0]	[0]	[0]	
Obs	232	1,008	280	1,520
LC	0.5482	0.3477	0.1954	0.3473
LE	0.9200	0.1346	0.0004	0.3243
GC	0.5541	0.1866	0.0107	0.1522
GE	0.8293	0.1203	0.0013	0.2028

Note: Top panel: Probability that estimate =0 is shown in brackets below estimate. Bottom panel displays estimated average values of the dependent variable for each treatment.

Table 16: Estimation. Errors. Overall and by type of agency.

In a fashion similar to the one used in section 4.1, several tests are shown in a schematic form in table 17 (the values of the tests can be found in table 19 in appendix B).

Dep. Var.:	ERRA	ERRB	ERRC	ERR
LC	+	+	+	+
LE	-	-		
GC	+	+		+
GE	-	-		-
LC=GC		GC	GC	GC
LE=GE		GE	LE	GE
LC=LE	LC	LE	LE	LE
GC=GE	GC	GE		

Note: Top panel: Empty if data fits predictions; "+" if observed errors are higher than predicted; "-" otherwise. Bottom panel: Empty if no difference, treatment with less errors otherwise.

Table 17: Errors tests. Predictions and inter-treatment comparisons.

The top panel tests the accuracy of predictions and shows that the data do not fit them. In particular, errors are usually higher than predicted in C treatments but lower than predicted in E ones. This is consistent with the Dominance results, which indicate that "too many" people evade when they should comply (C treatments) and comply when they should evade (E treatments). The main conclusion, thus, is basically the same as the one found for

Dominance in Result 1, and subject to the same qualifications.

The first two lines of the bottom panel are the important ones: they show the tests for the SCR hypothesis. Given the minimum variability in the extreme cases (when the agency is too soft, $q = A$, or too tough, $q = C$), the relevant tests are those for the medium one, and this one shows clearly that the *Global* treatments lead to less errors per capita than the *Lottery* ones. In other words, the SCR hypothesis is strongly supported.

Result 3 *Superiority of Contingent Rule (SCR): From the agency's perspective, the contingent rule is better than the cut-off one.*

The last two lines test whether there are significant differences between E and C treatments and show (again focusing on the medium case) that the first lead to less errors than the second. Again, this can be linked to the Dominance analysis, where E treatments show a higher degree of coincidence with predictions than C ones. This means, in other words, than in the latter many people evaded when they should have complied, and the higher number of associated errors thus explains the present result.

Finally, it is important to notice that all these findings are also supported graphically, as shown in figures 10 to 15. It can be clearly seen there that G treatments (i.e., those in which the contingent policy is implemented) lead to (weakly) less errors than L ones (those in which the cut-off one is used). It also shows the (expected) result that errors are a decreasing function of the agency's "toughness", which is consistent with GG's comparative statics.

Result 4 *Effect of agency's type (EAT): Errors decrease with the agency's "toughness".*

5 Conclusions

The empirical analysis of tax evasion is problematic because of the reluctance of both taxpayers and tax agencies to provide the relevant information. This study, therefore, uses experimental data as a second-best alternative and focuses on the testing of some of the theoretical predictions obtained in Sanchez Villalba, 2007), though the richness of the dataset also allows for the investigation of other interesting hypotheses related to decision-making processes and the global game theory.

Results are strongly supportive of the main prediction of the TEGG model, namely, that a tax agency using a contingent auditing policy would do better than if it used the standard

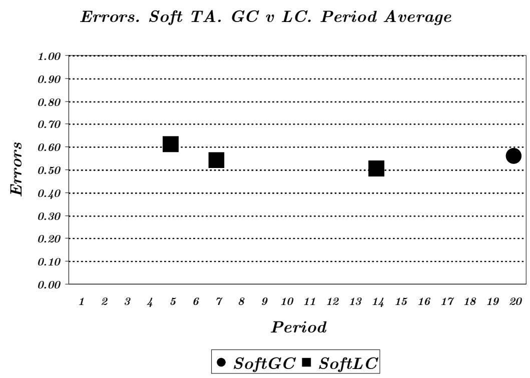


Figure 10: Errors. Soft agency. GC v LC.

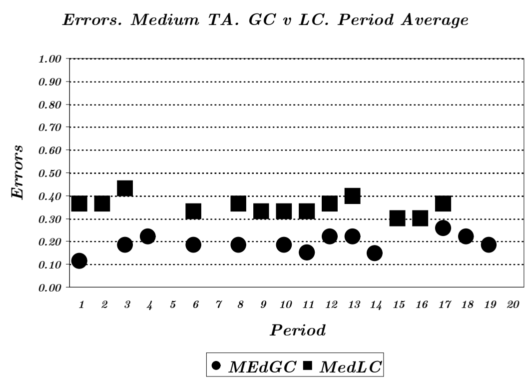


Figure 11: Errors. Medium agency. GC v LC.

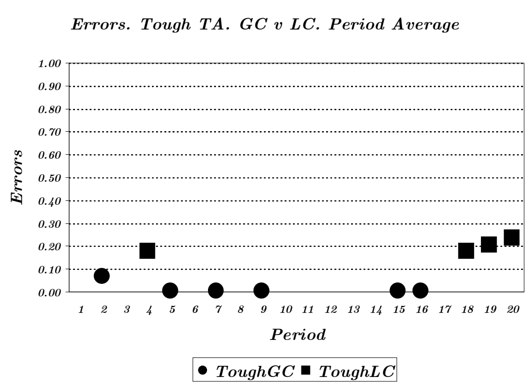


Figure 12: Errors. Tough agency. GC v LC.

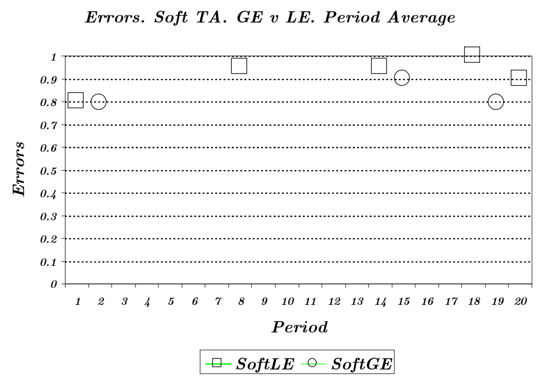


Figure 13: Errors. Soft agency. GE v LE.

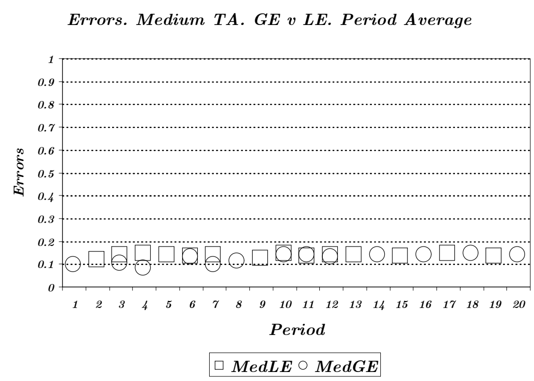


Figure 14: Errors. Medium agency. GE v LE.

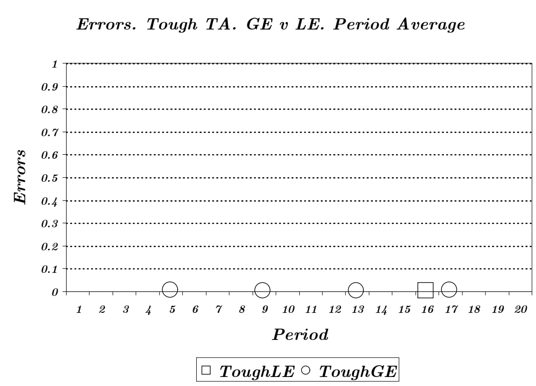


Figure 15: Errors. Tough agency. GE v LE.

cut-off one. The negative externality between taxpayers generated by the contingent policy and the associated strategic uncertainty it creates seem to be the powerful forces behind this result.

Also supported by the data are the predictions derived from the comparative statics of global games: evasion is higher in E treatments than in C ones, evasion is a decreasing function of signals, and errors decrease with agency's "toughness".

The picture, so encouraging in qualitative terms, is however radically different when considering it quantitatively: in general, the numerical predictions of the theory are rejected by the data. This is true for the medium cases (when the signal is medium), but not for the extreme ones though: in the latter, results are as expected and support the idea that people are, at least, "Rudimentary" and (intuitively) understand the concept of dominance in simple scenarios. Medium cases, on the other hand, show that most people do not use higher-order beliefs when making their decisions (not even in this simple experiment, in which only two iterations are needed). In spite of this, many times they do choose the actions predicted by the theory of global games, usually after playing the game a few times. This "learning" result is not so surprising, as it was already hinted by Carlsson and van Damme, 1993) and found experimentally by Cabrales *et al.*, 2002). Other factors also seem to affect decisions, like the tension between the risk-dominant and payoff-dominant equilibria, with their predicted effects closely mimicked by the data. More worrying, however, is the apparently pervasive presence of a significant group of people ("chance maximisers") who choose their strategies without taking into account all the available information (in this particular experiment, the payoffs in different scenarios) and that lead to the largest differences between observed and predicted actions (treatment LC). This concern is connected to the main result derived from the analysis of questionnaire data, which suggests that those with knowledge of game theory ("sophisticated" agents) are more likely to play according to predictions than those without that knowledge ("simple" agents). This can have an impact on policy-making, as one would expect higher degree of "sophistication" among firms than among individual taxpayers (though the latter group can change their status if they have access to sophisticated professional advice).

The bottom line is, therefore, that though people may not use higher-order beliefs, many times they end up choosing the same actions than the ones predicted by HOBs. Consequently, this ensures that predictions are usually supported in qualitative terms (comparative statics and inter-treatment comparisons) but rejected in quantitative ones. Nonetheless, the latter problem can be deemed as a minor one because of two mitigating factors: First, the discreteness of the model can work against it because it amplifies small differences and thus make the data-predictions matches more difficult (something already highlighted by Heinemann *et al.*, 2002)). And second, the parameters of the model were explicitly chosen to discourage said matches. Thus, the fact that the data does support (qualitatively) the predictions in these most demanding conditions makes one believe the theory would be an even better predictor in more favourable environments.

References

J. Alm and M. McKee. Tax compliance as a coordination game. *Journal of Economic Behavior and Organization*, 54:297–312, 2004.

A. Bosch-Domenech, J. G. Montalvo, R. Nagel, and A. Satorra. One, two, (three), infinity,...: Newspaper and lab beauty-contest experiments. *The American Economics Review*, 92(5):1687–1701, December 2002.

A. Cabrales, R. Nagel, and R. Armenter. Equlibrium selection through incomplete information in coordination games: An experimental study. *mimeo*, February 2002.

H. Carlsson and E. van Damme. Global games and equilibrium selection. *Econometrica*, 61(5):989–1018, 1993.

U. Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

J. C. Harsanyi and R. Selten. A general theory of equilibrium selection in games. *Cambridge, MA. MIT Press.*, 1988.

F. Heinemann, R. Nagel, and P. Ockenfels. Speculative attacks and financial architecture: Experimental analysis of coordination games with public and private information. *mimeo*, June 2002.

F. Heinemann, R. Nagel, and P. Ockenfels. Measuring strategic uncertainty in coordination games. *mimeo*, December 2004.

F. Heinemann, R. Nagel, and P. Ockenfels. The theory of global games on test: experimental analysis of coordination games with public and private information. *Econometrica*, 72(5):1583–1599, September 2004.

C. A. Holt and S. K. Laury. Risk aversion and incentive effects. *The American Economics Review*, 92(5):1644–1655, December 2002.

S. Morris and H. S. Shin. Social value of public information. *Mimeo*, 2002.

J. Reinganum and L. Wilde. A note on enforcement uncertainty and taxpayer compliance. *The Quarterly Journal of Economics*, 103(4):793–798, 1988.

M. A. Sanchez Villalba. Anti-evasion auditing policy in the presence of common income shocks. *mimeo*, 2007.

R. Selten, A. Sadrieh, and K. Abbink. Money does not induce risk neutral behavior, but binary lotteries do even worse. *Theory and Decision*, 46:211–249, 1999.

D. O. Stahl and P. W. Wilson. Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, 25:309–327, 1994.

A Instructions for treatment GC[27]

Introduction

First of all, thank you very much for taking part in this experiment. It is important to start by saying that, though part of a serious research programme, this experiment is NOT a test. There are no "right" or "wrong" answers.

How it works

Before we do anything, we have to run through a few ground rules and instructions. After that we will move to the experiment proper, where you will be asked to make decisions in a number of economic situations presented to you. Finally you will get paid: on top of a show-up fee of £5, you will get a sum of money that will depend on your performance in the situations mentioned before.

The experiment consists of 5 stages:

- Instructions
- Trial rounds
- Experiment rounds
- Questionnaire
- Payment

We will go through these in detail below.

Ground rules

For the experiment to work we need to run it according to fairly strict rules, but there are not too many:

- From now until the end of the experiment, please do not talk (it will not take long!)
- If there is something you need to ask about the way the experiment works just raise your hand -the experimenter will come to your desk.
- Please do not use the computer until you are told to.

The Six Stages

1 Instructions

[27]Instructions for the other treatments were similar to these ones, with the logical changes in rules and parameters needed in each case.

The experimenter will read out the instructions. If you have questions, this is the time to deal with them. Just raise your hand and the experimenter will answer them privately.

2 Short quiz

This is to ensure that you understand the instructions.

3 Trial rounds

The experiment is organised in a series of rounds. Each round is a period in which you interact –via the computer only– with the other participants and make decisions that determine the amount of money you will get at the end of the session.

As a warm-up you will first take part in 2 trial rounds. These trial rounds are identical to the experiment rounds in every respect with one exception: the effect on payment. Trial rounds do NOT affect your reward at the end of the experiment. They allow you to check out the interface and familiarise yourself with the screen tables, buttons and commands. They also allow you to make mistakes without losing money.

4 Experiment rounds

This is the real thing. What you do during these rounds will determine the total amount of money you will get.

The following "Frequently Asked Questions" will lead you through the basic mechanics of the rounds.

4.1. What is this all about?

Let us start by saying that the experiment will consist of 20 experiment rounds. In each one of them the computer will pair you up with one other participant. Each of the other participants in the room is equally likely to be paired up with you.

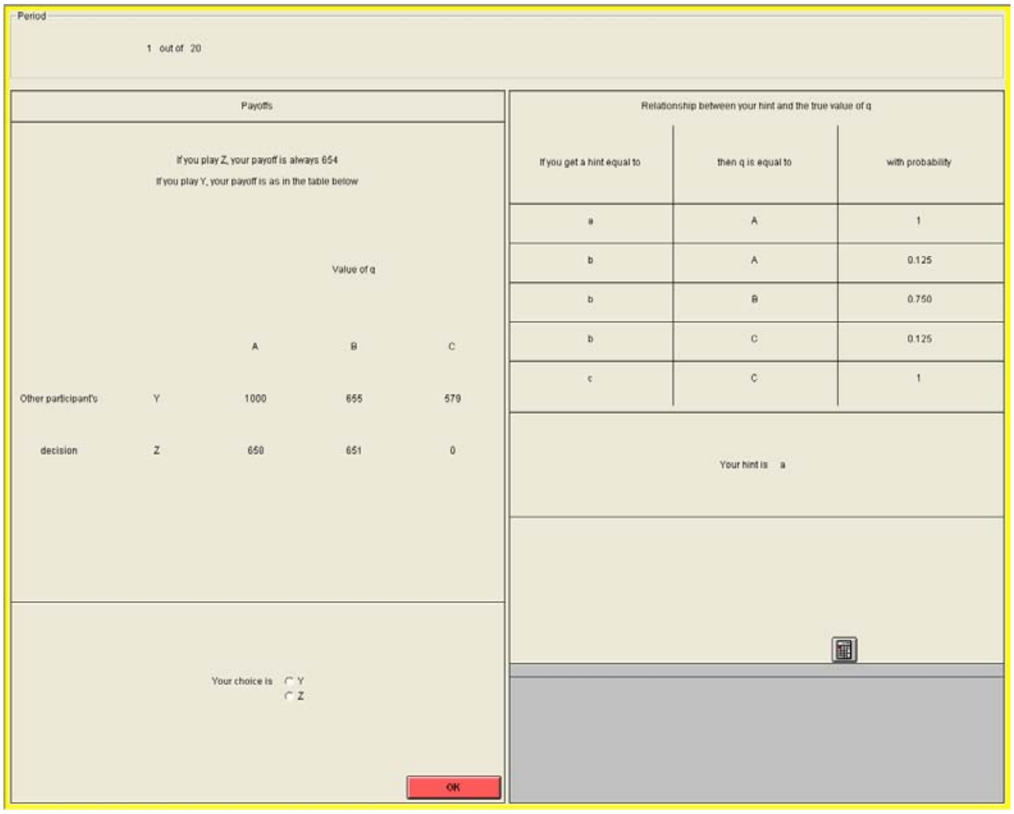
4.2. What do I have to do?

You have to choose one of two possible actions, namely Y or Z. You choose one or the other by clicking on your preferred option in the bottom left panel of the choice screen (see figure 1) and then pressing the "OK" button in the same panel.

Figure 1: Choice screen

4.3. How is my payoff for the round determined?

Your payoff for the round depends on your own action, the action of the other participant, and an unknown parameter called q.



4.4. But exactly how *is my payoff for the round determined?*

There are two cases to consider:

a. If you choose action Z, your payoff is 654 "experimental points" with certainty.

b. If you choose action Y, your payoff depends on both the value of q and the action of the other participant, as shown in the table below (and also in the top-left panel of the choice screen (see figure 1)):

		Value of q		
		A	B	C
Other participant's choice	Y	1000	655	579
	Z	658	651	0

That is, if you choose Z, you always get 654 "experimental points", regardless of what the other participant does and what the value of q is. But if you choose Y, then there are several cases to consider. Let us see some of them (remembering that in all of them you choose Y and your payoff is measured in "experimental points"):

If the other participant chooses Y and q equals A, then your payoff is 1000.

If the other participant chooses Y and q equals B, then your payoff is 655. And so on.

4.5. So how much money do I get then?

Your payoffs are transformed into money at a rate of: 1000 "experimental points" = 50 pence

That is, if your payoff for the round is, for example, 655 "experimental points", your corresponding money earnings are $655 \times 50/1000 = 32.75$pence.

Your session earnings are computed by adding up the money you got during the 20 experiment rounds.

4.6. But, what is q?

q is a parameter that can only take one of 3 values: A, B or C. In any given round, your computer will choose one of these 3 values, with probabilities 0.20, 0.60 and 0.20, respectively.

Intuitively, you can think of these probabilities in the following way: Consider an urn with 100 balls. 20 of them are labelled "A", 60 "B" and 20 "C". The value of q will be determined by the label of one of the 100 balls in the urn, chosen randomly (by the computer).

4.7. Is there anything I could use to make a more informed decision?

Yes, there is. Before you make a decision you will get a "hint". This hint will be known only to you and can only take one of 3 values: a, b or c. It provides some information about the value of the unknown parameter q, as shown in the following table (and in the top-right panel of the choice screen (figure 1)):

If hint is...	...then q is...	...with probability...
a	A	1.000
	A	0.125
b	B	0.750
	C	0.125
c	C	1.000

For any given round, your hint can be found immediately below this table in the choice screen (figure 1).

The table may seem a bit complicated but do not worry, it is not. It simply says that if your hint is equal to a, then you can be sure that q is equal to A. Analogously, if your hint is equal to c, then q is equal to C. When your hint is equal to b, however, you do not know for sure what the value of q is, but you can tell how likely each value is: q is equal to B with probability 0.750, while it is equal to A or C with probabilities 0.125 and 0.125, respectively.

Important note: Although q is the same for you and the other participant, your hints may differ from each other.

4.8. Anything else I should know before making my choice?

If you want to make some computations before choosing your action, you can press the calculator button on the choice screen (the small square button just above the darker area (see figure 1)). Pens and paper are available for those who prefer them: raise your hand and an experimenter will take them to your desk.

Also, it is worth mentioning that there is no "Back" button, so please make your decisions carefully and only press the "OK" or "Continue" buttons when you are sure you want to move to the next screen.

4.9. So I made my decision, what now?

After you submit your decision, you will be shown the action you chose and the payoff you got for the round, as well as the value that q took (see figure 2). By clicking on the "Continue" button you will move to a new round (if there is any still to be played).

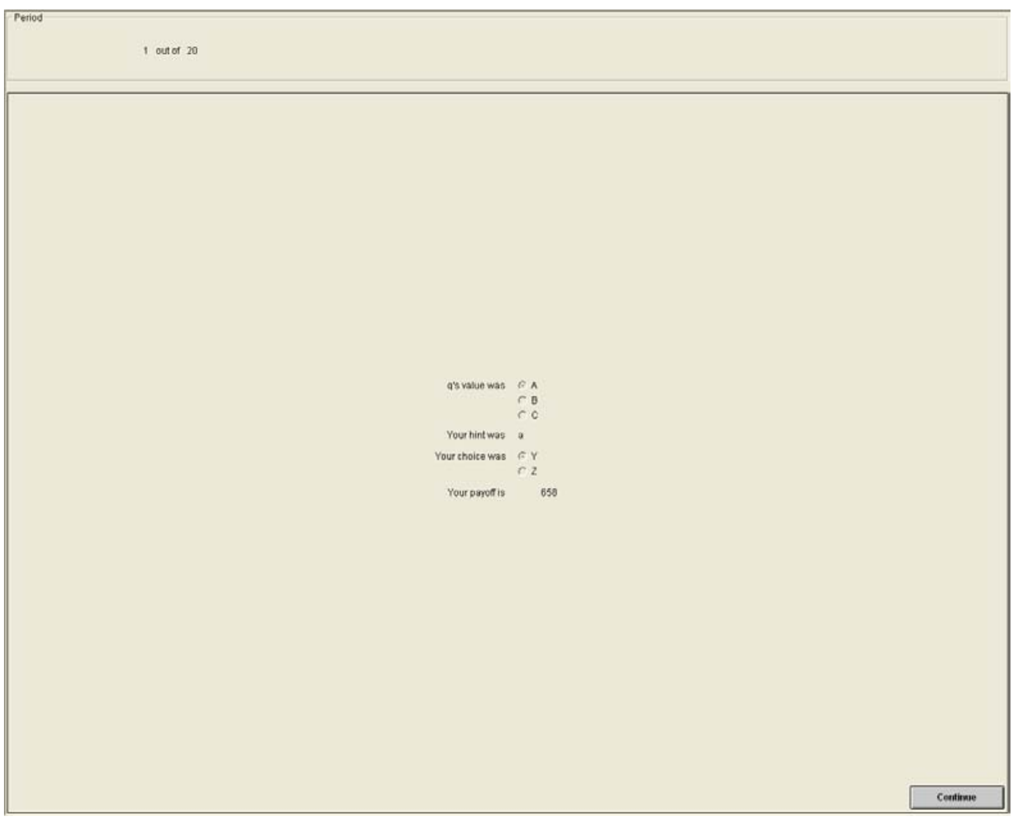


Figure 2: Results screen

4.10. And then? Is it the same over and over again?

Basically, yes. In every round, the structure is identical to the one described above: first a new q will be selected by the computer and you will be paired up with another participant, then you will be assigned a hint and will have to make a decision, and finally your payoff will be shown on the results screen.

You can check what happened in previous periods by taking a look at the darker area in the bottom-right panel of the choice screen (see figure 1). It includes information about the values adopted by q, the hints you got and the actions you chose in earlier rounds.

Important note: Every period is like a clean slate: the value of q, the participant you are paired up with and the hint you get may vary from round to round, but the RULES that determine them (explained in questions 4.6., 4.1. and 4.7.) do not. In short, rounds are independent: for example, you can think that in every round a new urn with 100 balls -20 "As", 60 "Bs" and 20 "Cs" - is used to determine the value of q, as explained in question 4.6. Similarly, the pairings and hints of a given round are independent of the pairings and hints of previous rounds.

5 Questionnaire

We will ask you a few questions that will help us to further understand the data collected in the session.

6 Payment

Finally! You will be paid a show-up fee of £5 plus the sum earned during the session, as explained in question 4.5.

And that is it. Once again, thank you very much for participating!

SHORT QUIZ

1. What is your payoff (in "experimental points") if you choose Y, the person paired-up with you chooses Z and q is equal to A? ....

2. What is your payoff (in "experimental points") if you choose Z, the person paired-up with you chooses Y and q is equal to C? ....

3. If your hint is equal to b, what is the probability that q is equal to A? ....

B Extra Tables

	DOMa	DOMb	DOMc	RD	AD	DOM
LC	.	0.0000	.	1.0000	0.0000	0.0000
LE	0.3231	0.0000	.	0.3180	0.0000	0.0000
GC	.	0.0000	0.1578	0.1552	0.0000	0.0000
GE	.	0.0000	0.0390	0.0418	0.0000	0.0000
LC=GC	.	0.0000	0.1578	0.1552	0.0000	0.0000
LE=GE	0.3231	0.0042	0.0390	0.1920	0.0042	0.0029
LC=LE	0.3231	0.0000	.	0.3180	0.0000	0.0000
GC=GE	.	0.0005	0.2170	0.4359	0.0005	0.0014

Note: Values of F-tests. Values below 5% imply the null hypothesis is rejected. Dots mean there is no variability in data as to compute the statistics.

Table 18: Dominance tests. Predictions and inter-treatment comparisons.

	ERRA	ERRB	ERRC	ERR
LC	0.3456	0.0000	0.0000	0.0518
LE	1.0000	0.1450	0.0004	0.3515
GC	0.2939	0.0000	0.0000	0.0147
GE	1.0000	0.145	0.0010	0.2445
LC	0.0000	0.0000	0.0000	0.0000
LE	0.0038	0.0000	1.0000	0.1441
GC	0.0092	0.0000	0.1575	0.0000
GE	0.0006	0.0000	0.2612	0.0092
LC=GC	0.9556	0.0000	0.0000	0.0000
LE=GE	0.1082	0.0000	0.0034	0.0000
LC=LE	0.0000	0.0000	0.0000	0.3555
GC=GE	0.0135	0.0000	0.2160	0.0112

Note: Top panel: Predicted values of dependent variable. Middle and bottom panels: Values of F-tests. Values below 5% imply the null hypothesis is rejected.

Table 19: Errors tests. Predictions and inter-treatment comparisons.

C Examples of categories

Expected-Payoff Maximisers (EPM): “If the hint was a, I selected action Y; otherwise, I selected action Z. There are only three outcomes that generate more than 654 points, and two of them only generate a negligible increase (relative to their risk). The only way to "gamble and win" is to play Y when the hint is b, and in that case, I am gambling that either my "opponent" has a hint of a (very unlikely), or my opponent has a hint of b, is risking that q is really A, and is right (also very unlikely). My risk is that my opponent plays Z, which is safer, and that q is B or C, which is likely. The risk/reward is far too high. When my hint is a or c the correct play is obvious - in the former case, playing Y always nets me more than 654, and in the latter, playing Y always nets me less than 654, no matter what my opponent does." (Subject #10, GC).

Chance Maximisers (CM): "If the hint is c, the best decision is always Z with a higher payoff. If the hint is b, it worths choosing Y, because thereis a probablility of 0.875 getting A or B, which are both higher than Z(654). If the hint is a, my decision is definitely Y." (Subject #20, LC).

Learners (L): "At first i played it safe and went with the guarantee button z and then i took more of a risk by chosing the y button every time i got the hint "a" or "b". because there was a higher probability of gaining more points." (Subject #18, GE).

Mixers/Experimenters (M/E): "If the hint was A, choice was Y. If the hint was C, choice was Z. If the hint was B, 80% of the time choice was Y and 20%, B." (Subject #15, LE).

Non-independent (NI): "If the hint came up as A i always selected choice Y as I would be better off (ie gaining more money) through doing so regardless of what the other participant chose. Conversely, if the value of q was C i always chose Z since I would be worse off if i choice Y despite what the other person selected. If the value of q came up as b i would go systematically throught the choices Y,Y,Z. This was my order since if q=b and q=a i would be better off selecting Y and if q=c i would be better off selecting Z. Since the probability of q=b was the highest i put Y at the beginning of the order. I used my knowledge of maths and probablities to calculate the order in which to place my choices." (Subject #2, GC).

Randomisers (R): "If the hint was a then i chose Y if the hint would have been c then i would have chosen Z. apart from this i just guessed randomly. the last 3 i thought i may as well take the risk as it was the end of the experiment." (Subject #19, LC).

Confused (C): "If the probability was lower than the other option, i chose the other option. I did not take risks in the cases where the probability could also go for the lowest amount. Becasue i dont know much about the probability theory so i decided to go for the safest method." (Subject #7, LC).

Risk-lovers (RL): "I chose Y every time unless I knew it was C. I was not given the hint a at any time. The difference between playing it safe and gambling with the Y option was small enough to make the experiment slightly more fun. I knew that I could lose 579, but only gain 421, but preferred the gamble." (Subject #7, GC).